Vazquez-Vilar, G., Guillén i Fàbregas, A. y Verdú, S. (2019). The Error Probability of Generalized Perfect Codes via the Meta-Converse. *IEEE Transactions on Information Theory,* 65(9), pp. 5705-5717.

DOI: https://doi.org/10.1109/TIT.2019.2906227

The Error Probability of Generalized Perfect Codes via the Meta-Converse

Gonzalo Vazquez-Vilar , *Member, IEEE*, Albert Guillén i Fàbregas , *Senior Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*— We introduce a definition of perfect and quasi-perfect codes for discrete symmetric channels based on the packing and covering properties of generalized spheres whose shape is tilted using an auxiliary probability measure. This notion generalizes previous definitions of perfect and quasi-perfect codes and encompasses maximum distance separable codes. The error probability of these codes, whenever they exist, is shown to coincide with the estimate provided by the meta-converse lower bound. We illustrate how the proposed definition naturally extends to cover almost-lossless source-channel coding and lossy compression.**

***Index Terms*— Shannon theory, perfect codes, quasi-perfect codes, maximum likelihood decoding, finite blocklength analysis, meta-converse, hypothesis testing, channel coding, joint source-channel coding, rate distortion theory.**

I. Introduction

In THE context of reliable communication, binary hypothesis testing has proved instrumental in the derivation of converse bounds to the error probability. Using this method, the sphere-packing bound on the channel coding reliability function was derived in [1] (see also [2]–[5] for alternative derivations and refinements). More recently, the meta-converse of Polyanskiy et al. [6, Th. 27] proved that a surrogate binary hypothesis test can be used to accurately lower bound the error probability in the finite blocklength regime. The non-Bayesian optimal performance of binary hypothesis testing between distributions P_0 and P_1 is characterized by the tradeoff $\alpha_\beta(P_0, P_1)$, where α denotes the smallest error under P_0 achievable by any test with error under P_1 at most β (we refer the reader to Section II for a formal definition). Then, [6, Th. 27] establishes the following lower bound on the error probability of a code $\mathcal{C}$ with cardinality M used over a channel $P_{Y|X}$,

$$P_{\mathrm{e}}(\mathcal{C}) \geq \inf_{P_X} \sup_{Q_Y} \left\{ \alpha_{\frac{1}{M}}\left(P_X \times P_{Y|X}, P_X \times Q_Y\right) \right\}. \tag{1}$$

This bound, or the more general [6, Th. 26], are sometimes referred to as meta-converse bounds, since many previous converse bounds in the literature can be proven as corollaries via relaxation. Particularized for n-uses of a memoryless binary symmetric channel (BSC), the meta-converse bound (1) recovers the sphere-packing bound for BSCs [7, eq. (5.8.19)] (see [6, Sec. III.H]). In this setting, the right-hand side of (1) coincides with the exact error probability whenever perfect or quasi-perfect codes exist. In particular, a binary code is said to be *perfect* if non-overlapping Hamming spheres of radius t centered on the codewords exactly fill out the space. Similarly, a *quasi-perfect* code is defined as a code in which Hamming spheres of radius t centered on the codewords are non-overlapping and Hamming spheres of radius $t+1$ cover the space, possibly with overlaps. This definition coincides with that of *sphere-packed codes* introduced by Gallager [7, Sec. 5.8]. Since quasi-perfect codes attain the lower bound (1), they achieve the minimum error probability in a BSC among all the codes with the same blocklength and rate.

In this work, we generalize the definition of perfect and quasi-perfect codes beyond Hamming distance and show their optimality for general discrete channels under certain symmetry conditions. The new definition, which is channel-dependent, follows from the packing and covering properties of generalized spheres whose shape is tilted using an auxiliary probability measure. We show that generalized perfect and quasi-perfect codes attain equality in (1). Therefore, they achieve the minimum error probability among all the codes with the same blocklength and rate. As an example, we study a family of q-ary symmetric erasure channels and we show that maximum-distance separable (MDS) codes are generalized quasi-perfect for these channels. As a result, we obtain an alternative proof of the optimality of MDS codes for q-ary symmetric erasure channels. Extensions to almost-lossless source-channel coding and lossy compression under an excess-distortion constraint are discussed.

This work was supported in part by the European Research Council (ERC) under Grant 714161 and Grant 725411, in part by the Spanish Ministry of Economy and Competitiveness under Grant TEC2016-78434-C3 and Grant IJCI-2015-27020, in part by the National Science Foundation under Grant CCF-1513915, in part by the Center for Science of Information, and in part by the NSF Science and Technology Center under Grant CCF-0939370. This paper was presented in part at the 2016 International Zürich Seminar and in part at the 2018 IEEE International Symposium on Information Theory.

G. Vazquez-Vilar is with the Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911 Leganés, Spain, and also with the Gregorio Marañón Health Research Institute, 28007 Madrid, Spain (e-mail: gvazquez@ieee.org).

A. Guillén i Fàbregas is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain, also with the Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

S. Verdú (e-mail: verdu@informationtheory.org).

Communicated by M. R. Bloch, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Our results are related to previous works. A tightened version of the meta-converse, derived for a fixed code, was shown to coincide with the exact error probability in [8, Th. 1]. In contrast to [8], in this paper we show that the bound (1), which applies to every code of cardinality M, also yields the exact error probability in certain cases. Hamada [9] also studied a generalization of perfect and quasi-perfect codes beyond Hamming distance. Using a variation of the Fano metric [10, eq. (9.10)], Hamada derived a lower bound to the channel coding error probability. Our definition of quasi-perfect codes includes [9, Definition 1] as a special case and recovers Hamada's condition for achieving minimum error probability [9, Th. 3]. Nevertheless, the class of codes considered here is more general than that in [9] and shows connections not previously treated in the literature.

The structure of the paper is as follows. In Section II we introduce the binary hypothesis testing framework and notation used in the rest of the paper. In Section III we introduce the system model and show the optimality of the so-called generalized quasi-perfect codes. A family of erasure channels is studied in detail under this formulation in Section IV and the optimality of MDS codes is shown. Sections V and VI extend the notion of generalized quasi-perfect codes to almost-lossless source-channel coding and lossy compression under maximum excess-distortion probability, respectively. Section VII closes the paper with some final remarks.

II. Binary Hypothesis Testing

Consider a non-Bayesian binary hypothesis test discriminating between distributions P_0 and P_1 defined over some discrete alphabet $\mathcal{Z}$.[1] Let $T(z) \in [0,1]$ denote the probability of the test deciding hypothesis 0 (corresponding to P_0) given an observation z. Then, $1-T(z)$ is the probability of deciding hypothesis 1 (i.e., P_1). Let $\pi_{j|i}(T)$ denote the probability that test T decides j when i is the true hypothesis, i.e.,

$$\pi_{0|1}(T) \triangleq \sum_z T(z) P_1(z), \tag{2}$$

$$\pi_{1|0}(T) \triangleq 1 - \sum_z T(z) P_0(z), \tag{3}$$

and we denote the minimum error probability $\pi_{1|0}$ among all tests T with $\pi_{0|1}$ at most β, as

$$\alpha_\beta\big(P_0, P_1\big) \triangleq \inf_{T:\pi_{0|1}(T)\le\beta} \pi_{1|0}(T). \tag{4}$$

Neyman and Pearson provided in [11] an explicit form for the test achieving the optimal tradeoff $\alpha_\beta\big(P_0, P_1\big)$. In particular, for any $\gamma \ge 0$, $\theta \in [0,1]$, an optimal test is given by

$$T_{\mathrm{NP}}(z) \triangleq \mathbb{1}\left[\frac{P_0(z)}{P_1(z)} > \gamma\right] + \theta\,\mathbb{1}\left[\frac{P_0(z)}{P_1(z)} = \gamma\right], \tag{5}$$

where $\mathbb{1}[\cdot]$ denotes the indicator function. T_{NP} achieves the optimal tradeoff $\alpha_\beta\big(P_0, P_1\big) = \pi_{1|0}(T_{\mathrm{NP}})$ when γ and θ are chosen such that $\beta = \pi_{0|1}(T_{\mathrm{NP}})$. The result is usually known as the Neyman-Pearson (NP) lemma. A direct consequence of the NP lemma is the following characterization of the optimal error probability tradeoff $\alpha_\beta\big(P_0, P_1\big)$.

Lemma 1: For any non-Bayesian binary hypothesis test discriminating between P_0 and P_1,

$$\alpha_\beta\big(P_0, P_1\big) = \sup_{\gamma\ge 0}\left\{\mathbb{P}\left[\frac{P_0(Z_0)}{P_1(Z_0)} \le \gamma\right] + \gamma\,\mathbb{P}\left[\frac{P_0(Z_1)}{P_1(Z_1)} > \gamma\right] - \gamma\beta\right\}, \tag{6}$$

where $Z_i \sim P_i$, $i = 0, 1$.

Proof: See Appendix A. ∎

III. Generalized Perfect Codes

An equiprobable message $m \in \{1, \dots, M\}$ is to be transmitted over a channel with transition probability $P_{Y|X}$, input $x \in \mathcal{X}$ and output $y \in \mathcal{Y}$, and where $\mathcal{X}$ and $\mathcal{Y}$ are the one-shot input/output discrete alphabets.[2] A channel code is the set of codewords $\mathcal{C} = \{x_1, \dots, x_M\}$, $x_i \in \mathcal{X}$ for $i = 1, \dots, M$, assigned to each of the messages. Under maximum likelihood (ML) decoding, the error probability for the code $\mathcal{C}$ is

$$P_{\mathrm{e}}(\mathcal{C}) = 1 - \frac{1}{M}\sum_y \max_{x\in\mathcal{C}} P_{Y|X}(y|x). \tag{7}$$

Henceforth, we will restrict attention to the following class of random transformations.

Definition 1: Let $F_x(\tau) \triangleq \mathbb{P}\big[P_{Y|X}(Y|x) \ge \tau\big]$, where $Y \sim P_{Y|X=x}$ and $\tau \in [0,1]$. A channel $P_{Y|X}$ is *symmetric* if $F_x(\tau)$ does not depend on the input x,

$$F_x(\tau) = F(\tau), \quad \forall x \in \mathcal{X}, \quad \tau \in [0,1]. \tag{8}$$

In the special case of discrete memoryless channels, Definition 1 implies that the rows of the channel transition matrix (with inputs as rows and outputs as columns), $P_{Y|X}(\cdot|x)$, are permutations of each other. This definition coincides with that of *uniformly dispersive* channels of Massey [12, Sec. 4.2] and is less restrictive than those of Cover and Thomas [13, Sec. 7.2] and Gallager [7, p. 94]. The definition in [13, Sec. 7.2] additionally requires that the columns of the channel transition matrix are permutations of each other. The definition in [7, p. 94] requires the channel transition matrix to be partitioned in submatrices such that each submatrix fulfills the conditions in [13, Sec. 7.2]. Relations among these notions are investigated in [14, Sec. VI.B].

Let Q be an auxiliary distribution defined on the output alphabet $\mathcal{Y}$. For an observation $y \in \mathcal{Y}$, the codeword $x \in \mathcal{C}$ that maximizes the metric $P_{Y|X}(y|x)$ also maximizes the metric $q(x,y) = \frac{P_{Y|X}(y|x)}{Q(y)}$. We conclude that the decoding regions induced by the ML decoder (with metric $P_{Y|X}(y|x)$) and those of the maximum metric decoder (with metric $q(x,y)$) coincide. This obvious fact proves to be instrumental next.

For any $\tau \ge 0$ and any auxiliary distribution Q defined over $\mathcal{Y}$, we define $\mathcal{S}_x(\tau, Q) \in \mathcal{Y}$ to be the set of outputs y

[1] The restriction to discrete alphabets can be avoided by simply replacing the ratio of probability mass functions by the corresponding Radon-Nykodim derivative.

[2] For example, for a BSC with crossover probability ϵ and blocklength n, $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$ and $P_{Y|X}(y|x) = \epsilon^{w(x\oplus y)}(1-\epsilon)^{n-w(x\oplus y)}$ where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ denote the channel input and output, respectively, and $w(\cdot)$ denotes the Hamming weight.

with likelihood given input x at least $\tau\, Q(y)$, i.e.,

$$\mathcal{S}_x(\tau, Q) \triangleq \left\{ y \in \mathcal{Y} \,\middle|\, \frac{P_{Y|X}(y|x)}{Q(y)} \geq \tau \right\}. \tag{9}$$

The shape of $\mathcal{S}_x(\tau, Q)$ in (9) is tilted via the auxiliary probability measure Q. Note that in contrast to Definition 1, $\mathcal{S}_x(\tau, Q)$ is defined for any $\tau \geq 0$, not necessarily $\tau \in [0, 1]$. We refer to $\mathcal{S}_x(\tau, Q)$ as a sphere of radius τ centered on x, although in general $\mathcal{X} \neq \mathcal{Y}$ and $q(x, y) \triangleq \frac{P_{Y|X}(y|x)}{Q(y)}$ need not be a distance measure. This metric is equivalent to the Fano metric [10, eq. (9.10)], defined as $-\log q(x, y) = \log \frac{Q(y)}{P_{Y|X}(y|x)}$. For channels such as the BSC, $\log P_{Y|X}(y|x)$ is an affine function of the Hamming distance between x and y and, hence, $\mathcal{S}_x(\tau, Q)$ becomes a Hamming sphere when the crossover probability is at most $\frac{1}{2}$ and Q is the equiprobable distribution.

The sphere $\mathcal{S}_x(\gamma, Q)$ corresponds to the decision region of the NP test (5) with $P_0 \leftarrow P_{Y|X}(\cdot|x)$, $P_1 \leftarrow Q(\cdot)$, and $\theta \leftarrow 1$. This motivates a new definition of perfect and quasi-perfect codes that will be presented next, and establishes the connection between these codes and the meta-converse bound (1). We define the interior and the outer shell of $\mathcal{S}_x(\tau, Q)$ as

$$\mathcal{S}_{\mathrm{i},x}(\tau, Q) \triangleq \left\{ y \in \mathcal{Y} \,\middle|\, \frac{P_{Y|X}(y|x)}{Q(y)} > \tau \right\}, \tag{10}$$

$$\mathcal{S}_{\mathrm{o},x}(\tau, Q) \triangleq \left\{ y \in \mathcal{Y} \,\middle|\, \frac{P_{Y|X}(y|x)}{Q(y)} = \tau \right\}. \tag{11}$$

We consider the set of distributions Q such that the tilted channel $\tilde{P}_{Y|X}(y|x) \propto \frac{P_{Y|X}(y|x)}{Q(y)}$ remains symmetric. More precisely, we define the set of symmetry-preserving auxiliary output distributions

$$\mathcal{Q} \triangleq \left\{ Q \in \mathcal{P}(\mathcal{Y}) \;\middle|\; F_x(\tau, Q) = F(\tau, Q),\ \forall x \in \mathcal{X},\ \tau \geq 0 \right\}, \tag{12}$$

where $F_x(\tau, Q) \triangleq \mathbb{P}\big[Y \in \mathcal{S}_x(\tau, Q)\big]$ with $Y \sim P_{Y|X=x}$, and $\mathcal{P}(\mathcal{A})$ denotes the set of all probability distributions over alphabet $\mathcal{A}$.

For symmetric channels $P_{Y|X}$, the set $\mathcal{Q}$ is non-empty as it always includes the equiprobable distribution, and it may include other auxiliary distributions. For example, consider a single use of the binary erasure channel (BEC) with erasure symbol e. In this case, any distribution of the form $Q(0) = Q(1) = \xi$, $Q(\mathsf{e}) = 1 - 2\xi$, does not alter the symmetry of the original channel, and therefore it is included in $\mathcal{Q}$. This example will be studied in detail in Section IV.

For a fixed $Q \in \mathcal{Q}$, we use the short-hand notation $\mathbb{Q}[\mathcal{A}] \triangleq \mathbb{P}[Y \in \mathcal{A}], Y \sim Q$.

Lemma 2: Let $P_{Y|X}$ be a symmetric channel and $Q \in \mathcal{Q}$ defined in (12). Then, the probabilities $\mathbb{Q}\big[\mathcal{S}_x(\tau, Q)\big]$, $\mathbb{Q}\big[\mathcal{S}_{\mathrm{i},x}(\tau, Q)\big]$ and $\mathbb{Q}\big[\mathcal{S}_{\mathrm{o},x}(\tau, Q)\big]$ are independent of $x \in \mathcal{X}$ for any $\tau \geq 0$.

Proof: We prove that the term $\mathbb{Q}\big[\mathcal{S}_{\mathrm{o},x}(\tau, Q)\big]$ does not depend on x. Then, the independence of the other two terms follows since

$$\mathbb{Q}\big[\mathcal{S}_x(\tau, Q)\big] = \sum_{\tau' \in \mathcal{L}_Q \cap [\tau, \infty)} \mathbb{Q}\big[\mathcal{S}_{\mathrm{o},x}(\tau', Q)\big], \tag{13}$$

$$\mathbb{Q}\big[\mathcal{S}_{\mathrm{i},x}(\tau, Q)\big] = \sum_{\tau' \in \mathcal{L}_Q \cap (\tau, \infty)} \mathbb{Q}\big[\mathcal{S}_{\mathrm{o},x}(\tau', Q)\big], \tag{14}$$

where the countable set $\mathcal{L}_Q$ is defined as

$$\mathcal{L}_Q \triangleq \left\{ \nu \in \mathbb{R} \,\middle|\, \exists x \in \mathcal{X}, \exists y \in \mathcal{Y},\ \frac{P_{Y|X}(y|x)}{Q(y)} = \nu \right\}. \tag{15}$$

To show that $\mathbb{Q}\big[\mathcal{S}_{\mathrm{o},x}(\tau, Q)\big]$ is independent of x, we write

$$\mathbb{Q}\big[\mathcal{S}_{\mathrm{o},x}(\tau, Q)\big] = \sum\nolimits_y Q(y) \mathbb{1}\left[P_{Y|X}(y|x) = \tau\, Q(y)\right] \tag{16}$$

$$= \frac{1}{\tau} \sum\nolimits_y P_{Y|X}(y|x) \mathbb{1}\left[P_{Y|X}(y|x) = \tau\, Q(y)\right] \tag{17}$$

$$= \lim_{\delta \to 0} \frac{1}{\tau} \left(F_x(\tau, Q) - F_x(\tau + \delta, Q)\right), \tag{18}$$

where (18) holds for any $Q \in \mathcal{Q}$ in view of (12). The result follows since $F_x(\tau, Q)$ does not depend on x for any $Q \in \mathcal{Q}$. ∎

Then, according to Lemma 2, we define for symmetric channels the probability measures

$$\mathsf{Q}(\tau) \triangleq \mathbb{Q}\big[\mathcal{S}_x(\tau, Q)\big], \tag{19a}$$

$$\mathsf{Q}_{\mathrm{i}}(\tau) \triangleq \mathbb{Q}\big[\mathcal{S}_{\mathrm{i},x}(\tau, Q)\big], \tag{19b}$$

$$\mathsf{Q}_{\mathrm{o}}(\tau) \triangleq \mathbb{Q}\big[\mathcal{S}_{\mathrm{o},x}(\tau, Q)\big]. \tag{19c}$$

For a fixed code $\mathcal{C}$ and auxiliary distribution $Q \in \mathcal{Q}$, we let $\eta \geq 0$ be the largest value such that $\bigcup_{x \in \mathcal{C}} \mathcal{S}_x(\eta, Q) = \mathcal{Y}$. Similarly, let $\nu \geq 0$ be the smallest value such that the codeword centered sets $\left\{\mathcal{S}_{\mathrm{i},x}(\nu, Q)\right\}_{x \in \mathcal{C}}$ are disjoint. We respectively refer to η and ν as the *covering* and *packing radii* of the code $\mathcal{C}$ with respect to Q. Intuitively, $\mathcal{S}_{\mathrm{i},x}(\nu, Q)$ is the largest sphere packed inside the ML decoding region corresponding to $x \in \mathcal{C}$. Similarly, $\mathcal{S}_x(\eta, Q)$ is the smallest sphere centered at $x \in \mathcal{C}$ which completely covers the corresponding ML decoding region.

Definition 2: A code $\mathcal{C}$ is *generalized perfect* for $P_{Y|X}$, if there exists $\gamma \geq 0$ and $Q \in \mathcal{Q}$ such that the output alphabet can be partitioned into the codeword-centered sets $\mathcal{S}_x(\gamma, Q)$, i.e.,

$$\bigcup_{x \in \mathcal{C}} \mathcal{S}_x(\gamma, Q) = \mathcal{Y} \tag{20}$$

where the union is disjoint. A code is *generalized quasi-perfect* if there exists $\gamma \geq 0$ and $Q \in \mathcal{Q}$ such that (20) is satisfied and the codeword-centered sets $\left\{\mathcal{S}_{\mathrm{i},x}(\gamma, Q)\right\}_{x \in \mathcal{C}}$ are disjoint.[3,4]

Note that for generalized quasi-perfect codes the covering and packing radii coincide. The definition of quasi-perfect

[3] Occasionally, it is convenient to specify the auxiliary output distribution under which the code is generalized perfect or quasi-perfect, in which case we refer to the code as generalized Q perfect/quasi-perfect.

[4] While the sets $\mathcal{S}_x(\gamma, Q)$ and $\mathcal{S}_{\mathrm{i},x}(\gamma, Q)$ are a function of the parameters γ and Q, they depend only on their product (see (9) and (10)). Therefore, the two parameters $\gamma \geq 0$ and $Q \in \mathcal{Q}$ appearing in the definition of generalized perfect and quasi-perfect codes can be replaced by a single unnormalized measure $f(y) = \gamma\, Q(y)$.

codes includes perfect codes as a special case. To avoid ambiguities, for perfect codes we require that γ is the largest value satisfying (20). For this value of γ,

$$\bigcup_{x\in\mathcal{C}} \mathcal{S}_{\mathrm{i},x}(\gamma, Q) \subset \mathcal{Y}. \tag{21}$$

The main result in this work, Theorem 1, is a consequence of the following converse result, which is a refinement of [10, (9.15)–(9.16)].

Lemma 3: Let $P_{Y|X}$ be a symmetric channel and let $Q \in \mathcal{Q}$. The error probability of any code $\mathcal{C}$ with M codewords satisfies, for any $\gamma \geq 0$ and any $Q \in \mathcal{Q}$,

$$P_{\mathrm{e}}(\mathcal{C}) \geq \gamma\left(\mathsf{Q}_{\mathrm{i}}(\gamma) - \frac{1}{M}\right) + \sum_{\tau\in\mathcal{L}_Q\cap[0,\gamma]} \tau\mathsf{Q}_{\mathrm{o}}(\tau), \tag{22}$$

where $\mathcal{L}_Q$ is defined in (15). Furthermore, the lower bound (22) holds with equality if and only if $\mathcal{C}$ is generalized quasi-perfect and γ and Q are those parameters (not necessarily unique) satisfying the conditions in Definition 2.

Proof: Let $\mathcal{C} = \{x_1, \ldots, x_M\}$ be an arbitrary code. We consider a deterministic ML decoder which partitions the output space into disjoint decoding regions $\{\mathcal{D}_1, \ldots, \mathcal{D}_M\}$. The error probability (7) becomes

$$P_{\mathrm{e}}(\mathcal{C}) = 1 - \frac{1}{M}\sum_{m=1}^{M}\sum_{y\in\mathcal{D}_m} P_{Y|X}(y|x_m). \tag{23}$$

For an observed y, the codeword $x \in \mathcal{C}$ that maximizes the metric $P_{Y|X}(y|x)$ coincides with the one maximizing the metric $q(x, y) = \frac{P_{Y|X}(y|x)}{Q(y)}$. Then, using the definition of the covering and packing radii η and ν, respectively, it follows that

$$\mathcal{S}_{\mathrm{i},x_m}(\nu, Q) \subseteq \mathcal{D}_m \subseteq \mathcal{S}_{x_m}(\eta, Q), \tag{24}$$

for $1 \leq m \leq M$. As a result, $\mathcal{D}_m$ can be decomposed as

$$\mathcal{D}_m = \mathcal{S}_{\mathrm{i},x_m}(\nu, Q) \bigcup \left(\bigcup_{\tau\in\mathcal{L}_Q\cap[\eta,\nu]} \left(\mathcal{D}_m \cap \mathcal{S}_{\mathrm{o},x_m}(\tau, Q)\right)\right), \tag{25}$$

and (23) becomes

$$P_{\mathrm{e}}(\mathcal{C}) = 1 - \frac{1}{M}\sum_{m=1}^{M}\Bigg(\sum_{y\in\mathcal{S}_{\mathrm{i},x_m}(\nu,Q)} P_{Y|X}(y|x_m) + \sum_{\tau\in\mathcal{L}_Q\cap[\eta,\nu]}\sum_{y\in\mathcal{D}_m\cap\mathcal{S}_{\mathrm{o},x_m}(\tau,Q)} P_{Y|X}(y|x_m)\Bigg). \tag{26}$$

Since $\frac{P_{Y|X}(y|x)}{Q(y)} = \tau$ for any $y \in \mathcal{S}_{\mathrm{o},x}(\tau, Q)$, we write

$$\sum_{y\in\mathcal{S}_{\mathrm{i},x}(\nu)} P_{Y|X}(y|x) = \sum_{y\in\mathcal{S}_{\mathrm{i},x}(\nu,Q)} \frac{P_{Y|X}(y|x)}{Q(y)} Q(y) \tag{27}$$

$$= \sum_{\tau\in\mathcal{L}_Q\cap(\nu,\infty)}\sum_{y\in\mathcal{S}_{\mathrm{o},x}(\tau,Q)} \tau Q(y) \tag{28}$$

$$= \sum_{\tau\in\mathcal{L}_Q\cap(\nu,\infty)} \tau\mathsf{Q}_{\mathrm{o}}(\tau), \tag{29}$$

where in (29) we used Lemma 2 and $\mathsf{Q}_{\mathrm{o}}(\tau) = \mathbb{Q}\left[\mathcal{S}_{\mathrm{o},x}(\tau, Q)\right]$ as defined in (19c). Similarly,

$$\sum_{y\in\mathcal{D}_m\cap\mathcal{S}_{\mathrm{o},x}(\tau,Q)} P_{Y|X}(y|x) = \sum_{y\in\mathcal{D}_m\cap\mathcal{S}_{\mathrm{o},x}(\tau,Q)} \tau Q(y) \tag{30}$$

$$= \tau\mathsf{Q}_{\mathrm{o},m}(\tau) \tag{31}$$

where we abbreviate $\mathsf{Q}_{\mathrm{o},m}(\tau) \triangleq \mathbb{Q}\left[\mathcal{D}_m \cap \mathcal{S}_{\mathrm{o},x_m}(\tau, Q)\right]$.

Substituting (29) and (31) in (26), yields

$$P_{\mathrm{e}}(\mathcal{C}) = 1 - \Bigg(\sum_{\tau\in\mathcal{L}_Q\cap(\nu,\infty)} \tau\mathsf{Q}_{\mathrm{o}}(\tau) + \frac{1}{M}\sum_{m=1}^{M}\sum_{\tau\in\mathcal{L}_Q\cap[\eta,\nu]} \tau\mathsf{Q}_{\mathrm{o},m}(\tau)\Bigg). \tag{32}$$

Since $\{\mathcal{D}_m\}_{m=1}^{M}$ defines a partition of the output space, $\sum_{m=1}^{M}\mathbb{Q}\left[\mathcal{D}_m\right] = 1$. Using (25) and the definitions of $\mathsf{Q}_{\mathrm{i}}(\cdot)$ and $\mathsf{Q}_{\mathrm{o},m}(\cdot)$, we obtain

$$1 = \sum_{m=1}^{M}\mathbb{Q}\left[\mathcal{D}_m\right] = M\mathsf{Q}_{\mathrm{i}}(\nu) + \sum_{m=1}^{M}\sum_{\tau\in\mathcal{L}_Q\cap[\eta,\nu]} \mathsf{Q}_{\mathrm{o},m}(\tau). \tag{33}$$

Upon rearranging terms, (33) yields

$$\nu\left(\frac{1}{M} - \mathsf{Q}_{\mathrm{i}}(\nu)\right) = \frac{1}{M}\sum_{m=1}^{M}\sum_{\tau\in\mathcal{L}_Q\cap[\eta,\nu]} \nu\mathsf{Q}_{\mathrm{o},m}(\tau) \tag{34}$$

$$\geq \frac{1}{M}\sum_{m=1}^{M}\sum_{\tau\in\mathcal{L}_Q\cap[\eta,\nu]} \tau\mathsf{Q}_{\mathrm{o},m}(\tau). \tag{35}$$

Then, using (34)-(35) in (32), it follows that $P_{\mathrm{e}}(\mathcal{C}) \geq \Gamma(\nu)$ where

$$\Gamma(\nu) \triangleq 1 - \left(\sum_{\tau\in\mathcal{L}_Q\cap(\nu,\infty)} \tau\mathsf{Q}_{\mathrm{o}}(\tau) + \nu\left(\frac{1}{M} - \mathsf{Q}_{\mathrm{i}}(\nu)\right)\right). \tag{36}$$

For quasi-perfect codes satisfying Definition 2, there exist $Q \in \mathcal{Q}$ and $\gamma = \nu = \eta$ such that covering and packing radii coincide. Then, for this choice of parameters, inequality (35) becomes an equality and $P_{\mathrm{e}}(\mathcal{C}) = \Gamma(\gamma)$. We conclude that, for a generalized quasi-perfect code $\mathcal{C}$, (22) holds with equality for any choice (not necessarily unique) of γ and Q satisfying the conditions in Definition 2.

If $\mathcal{C}$ is not generalized quasi-perfect, $\nu > \eta$ for every $Q \in \mathcal{Q}$ and the inequality (35) is strict. Then, $P_{\mathrm{e}}(\mathcal{C}) > \Gamma(\nu)$. We next show that $P_{\mathrm{e}}(\mathcal{C}) > \Gamma(\gamma)$ for any choice of $\gamma \geq 0$ not necessarily equal to the packing radius ν. First, note that for $\gamma > \nu$, both (32) and (34)-(35) still hold substituting ν by γ. Then, the discussion above still applies.

Assume now that $\eta \leq \gamma < \nu$. We rewrite (32) as

$$P_{\mathrm{e}}(\mathcal{C}) = 1 + \frac{1}{M}\sum_{m=1}^{M}\sum_{\tau\in\mathcal{L}_Q\cap(\gamma,\nu]} \tau\,\Delta_m(\tau) - \sum_{\tau\in\mathcal{L}_Q\cap(\gamma,\infty)} \tau\mathsf{Q}_{\mathrm{o}}(\tau) - \frac{1}{M}\sum_{m=1}^{M}\sum_{\tau\in\mathcal{L}_Q\cap[\eta,\gamma]} \tau\mathsf{Q}_{\mathrm{o},m}(\tau). \tag{37}$$

where $\Delta_m(\tau) \triangleq \mathsf{Q}_{\mathrm{o}}(\tau) - \mathsf{Q}_{\mathrm{o},m}(\tau)$. Similarly, (33) becomes

$$1 = M\mathsf{Q}_{\mathrm{i}}(\gamma) + \sum_{m=1}^{M} \sum_{\tau \in \mathcal{L}_Q \cap [\eta,\gamma]} \mathsf{Q}_{\mathrm{o},m}(\tau) - \sum_{m=1}^{M} \sum_{\tau \in \mathcal{L}_Q \cap (\gamma,\nu]} \Delta_m(\tau). \tag{38}$$

Following analogous steps as in (34)-(35), via (37) we obtain

$$P_{\mathrm{e}}(\mathcal{C}) \geq \Gamma(\gamma) + \frac{1}{M} \sum_{m=1}^{M} \sum_{\tau \in \mathcal{L}_Q \cap (\gamma,\nu]} (\tau - \gamma)\Delta_m(\tau). \tag{39}$$

All terms in the inner sum in (39) satisfy $\tau - \gamma > 0$ and $\Delta_m(\tau) \geq 0$. If the code $\mathcal{C}$ is not generalized quasi-perfect, then, either $P_{\mathrm{e}}(\mathcal{C}) > \Gamma(\gamma)$ or $\Delta_m(\tau) > 0$ for at least one term in the sum. As the same proof steps follow for $\gamma < \eta$, we conclude that $P_{\mathrm{e}}(\mathcal{C}) > \Gamma(\gamma)$ for any $\gamma \geq 0$, $Q \in \mathcal{Q}$, provided that $\mathcal{C}$ is not quasi-perfect. ■

We are now ready to state the main result of this section, which shows that the ML decoding error probability of generalized perfect and quasi-perfect codes coincides with the meta-converse lower bound (1).

Theorem 1: Let $P_{Y|X}$ be a symmetric discrete channel and $\mathcal{C}$ be generalized quasi-perfect code. Then, $\mathcal{C}$ attains the minimum error probability among all codes with M codewords, which is given by

$$P_{\mathrm{e}}(\mathcal{C}) = \min_{P_X} \max_{Q \in \mathcal{Q}} \alpha_{\frac{1}{M}}\left(P_X \times P_{Y|X}, P_X \times Q\right) \tag{40}$$

$$= \max_{Q \in \mathcal{Q}} \alpha_{\frac{1}{M}}\left(P_{Y|X=x}, Q\right), \text{ for every } x \in \mathcal{X}. \tag{41}$$

Conversely, any code for which (40)-(41) hold is generalized quasi-perfect.

Proof: Let us consider the hypothesis test in (40). We apply Lemma 1 with $P_0 \leftarrow P_X \times P_{Y|X}$ and $P_1 \leftarrow P_X \times Q$. Using the definition of the set $\mathcal{S}_{\mathrm{i},x}(\cdot)$ and $\mathsf{Q}_{\mathrm{i}}(\cdot)$ in Lemma 2 yields

$$\alpha_{\frac{1}{M}}\left(P_X \times P_{Y|X}, P_X \times Q\right) = \sup_{\gamma \geq 0}\left\{ \sum_{x, y \notin \mathcal{S}_{\mathrm{i},x}(\gamma,Q)} P_X(x)P_{Y|X}(y|x) + \gamma\, \mathsf{Q}_{\mathrm{i}}(\gamma) - \frac{\gamma}{M} \right\}. \tag{42}$$

For any $y \in \mathcal{S}_{\mathrm{o},x}(\tau, Q)$, $\tau \in \mathcal{L}_Q$, where $\mathcal{L}_Q$ is defined in (15), it holds that $\frac{P_{Y|X}(y|x)}{Q(y)} = \tau$. Then,

$$\sum_{y \notin \mathcal{S}_{\mathrm{i},x}(\gamma,Q)} P_{Y|X}(y|x) = \sum_{\tau \in \mathcal{L}_Q \cap [0,\gamma]} \sum_{y \in \mathcal{S}_{\mathrm{o},x}(\tau,Q)} P_{Y|X}(y|x) \tag{43}$$

$$= \sum_{\tau \in \mathcal{L}_Q \cap [0,\gamma]} \sum_{y \in \mathcal{S}_{\mathrm{o},x}(\tau,Q)} \tau\, Q(y) \tag{44}$$

$$= \sum_{\tau \in \mathcal{L}_Q \cap [0,\gamma]} \tau\, \mathsf{Q}_{\mathrm{o}}(\tau), \tag{45}$$

which does not depend on x (see Lemma 2). Then, (42) becomes

$$\alpha_{\frac{1}{M}}\left(P_X \times P_{Y|X}, P_X \times Q\right) = \max_{\gamma \geq 0}\left\{ \sum_{\tau \in \mathcal{L}_Q \cap [0,\gamma]} \tau\, \mathsf{Q}_{\mathrm{o}}(\tau) + \gamma\, \mathsf{Q}_{\mathrm{i}}(\gamma) - \tfrac{\gamma}{M} \right\}. \tag{46}$$

According to (1), the right-hand side of (46) is a lower bound to $P_{\mathrm{e}}(\mathcal{C})$. According to Lemma 3, the term in braces in (46) is precisely the error probability of a generalized quasi-perfect code with parameters Q and γ. Therefore, whenever such a code exists the lower bound (46) is achievable and (40) holds with equality. Moreover, (41) holds since (46) does not depend on P_X for symmetric channels and $Q \in \mathcal{Q}$.

Let now $Q \in \mathcal{Q}$ achieve (40)-(41), and let γ be the maximizer in (46). It follows from Lemma 3 that the term in braces in (46) is the error probability of a code $\mathcal{C}$ if and only if $\mathcal{C}$ is generalized quasi-perfect and the parameters γ and Q satisfy the conditions in Definition 2. We conclude that, if (40)-(41) hold, $\mathcal{C}$ must be generalized quasi-perfect. ■

For any codebook $\mathcal{C} = \{x_1, \ldots, x_M\}$, we let $P_X^{\mathcal{C}}$ denote the empirical input distribution induced by $\mathcal{C}$, i. e., $P_X^{\mathcal{C}}(x) \triangleq \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}\{x = x_m\}$. It was shown in [8, Th. 1] that the error probability of any code can be expressed as

$$P_{\mathrm{e}}(\mathcal{C}) = \max_{Q}\left\{\alpha_{\frac{1}{M}}\left(P_X^{\mathcal{C}} \times P_{Y|X}, P_X^{\mathcal{C}} \times Q\right)\right\} \tag{47}$$

$$\geq \min_{P_X} \max_{Q}\left\{\alpha_{\frac{1}{M}}\left(P_X \times P_{Y|X}, P_X \times Q\right)\right\}, \tag{48}$$

Eq. (47) shows that the meta-converse bound, when applied to a fixed code $\mathcal{C}$, coincides with the exact error probability $P_{\mathrm{e}}(\mathcal{C})$. Theorem 1 shows that, under certain symmetry conditions, the relaxation (48) also coincides with the exact error probability, provided that a quasi-perfect code of cardinality M exists for this channel. Note that Theorem 1 is more general than the result obtained by Hamada in [9, Th. 3]. For instance, Theorem 1 can be used to prove the finite-blocklength optimality of MDS codes for q-ary erasure channels, as we show in the next section.

IV. Symmetric Erasure/Error Channels

Consider a symmetric erasure/error channel $P_{Y|X}$ with discrete input alphabet $\mathcal{X}$, $|\mathcal{X}| = q$, and output alphabet $\mathcal{Y} = \mathcal{X} \cup \{\mathsf{e}\}$ where e corresponds to the erasure symbol:

$$P_{Y|X}(y|x) = \begin{cases} 1-\delta-\epsilon, & y = x, \\ \delta, & y = \mathsf{e}, \\ \frac{\epsilon}{q-1}, & \text{otherwise.} \end{cases} \tag{49}$$

When $q = 2$, this channel includes as particular cases the BSC and the BEC by letting $\delta = 0$ and $\epsilon = 0$, respectively.

We consider n uses of this channel. Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$ denote the channel input and output, respectively. For a given pair of $\boldsymbol{x}$ and $\boldsymbol{y}$, we define the number of erasures and the number of flip-errors, respectively, as

$$e_{\boldsymbol{y}} \triangleq \sum_{i=1}^{n} \mathbb{1}[y_i = \mathsf{e}], \tag{50}$$

$$d_{x,y} \triangleq \sum_{i=1}^{n} \mathbb{1}[x_i \neq y_i \neq \mathsf{e}]. \tag{51}$$

The n-dimensional channel transition probability is given by

$$P_{Y|X}(y|x) = \delta^{e_y} \left(\tfrac{\epsilon}{q-1}\right)^{d_{x,y}} (1-\delta-\epsilon)^{n-e_y-d_{x,y}}. \tag{52}$$

We assume that $\frac{\epsilon}{q-1} < 1-\delta-\epsilon$. Otherwise, observing the transmitted symbol at the output of the channel would be less likely than observing any of the other $q-1$ symbols. Particularized to the BSC (with $q=2$, $\delta=0$), this assumption boils down to the crossover probability being $\epsilon < \frac{1}{2}$.

We define the auxiliary distribution

$$Q_Y^\star(y) \triangleq \tfrac{1}{c} \delta^{e_y} \left(\tfrac{\epsilon}{q-1}\right)^{\Psi(e_y)} (1-\delta-\epsilon)^{n-e_y-\Psi(e_y)}, \tag{53}$$

where c is a normalizing constant, and $\Psi(e) \geq 0$ is some non-negative function of the number of erasures e, which can be optimized over. Intuitively, $\Psi(e)$ corresponds to the average number of flip-errors that a good code can correct when the output sequence is affected by e erasures. For binary-input channels, a good choice for $\Psi(e)$ is given by

$$\Psi(e) = \max\left(0, \left\lfloor \tfrac{\lceil n - \log_2 M \rceil - e + 1}{2} \right\rfloor\right). \tag{54}$$

Since $Q_Y^\star(y)$ only depends on y via the number of erasures e_y, it does not affect the symmetry of the vector channel $P_{Y|X}$ and thus $Q_Y^\star \in \mathcal{Q}$. We particularize Theorem 1 for this channel and fix the tilting probability measure $Q = Q_Y^\star$ to obtain the following lower bound, which can be maximized over auxiliary functions $\Psi(e) \geq 0$.

Corollary 1: The error probability of any code $\mathcal{C}$ with cardinality M used over the channel (52) satisfies

$$P_{\mathrm{e}}(\mathcal{C}) \geq \sum_{e=0}^{n} \sum_{d=0}^{n-e} \binom{n}{e} \binom{n-e}{d} (q-1)^d \delta^e (1-\delta-\epsilon)^{n-e} \times \left(\varphi^{\max(d,\Psi(e))} - \frac{\varphi^{\Psi(e)}}{M}\right), \tag{55}$$

where $\varphi \triangleq \frac{\epsilon}{q-1}(1-\delta-\epsilon)^{-1} < 1$ and $\Psi(e) \geq 0$ is any positive function of the number of erasures e. Moreover, if $\mathcal{C}$ is a generalized quasi-perfect code that satisfies Definition 2 with $\gamma = c$ and $Q = Q_Y^\star$ then (55) holds with equality.

Proof: Let us consider the lower bound that follows from (46) by fixing $Q = Q_Y^\star$, defined in (53), and fixing $\gamma = c$ to be the normalization factor appearing in (53). In view of the channel symmetry and the choice of Q, we can write for any $x \in \mathcal{X}^n$

$$\max_{Q \in \mathcal{Q}} \left\{ \alpha_{\frac{1}{M}} \left(P_X \times P_{Y|X}, P_X \times Q\right)\right\} \geq \sum_{y \notin \mathcal{S}_{\mathrm{i},x}(c, Q_Y^\star)} P_{Y|X}(y|x) - c\left(\frac{1}{M} - \sum_{y \in \mathcal{S}_{\mathrm{i},x}(c, Q_Y^\star)} Q_Y^\star(y)\right). \tag{56}$$

For the choice $\gamma = c$ and $Q = Q_Y^\star$, the sets $\mathcal{S}_{\mathrm{i},x}(\gamma, Q)$ become

$$\mathcal{S}_{\mathrm{i},x}(c, Q_Y^\star) = \left\{ y \in \mathcal{Y} \,\middle|\, d_{x,y} < \Psi(e_y) \right\}. \tag{57}$$

We parametrize each output sequence y by the indices $e = e_y \in [0, n]$ and $d = d_{x,y} \in [0, n - e_y]$. For a given x, there are exactly $\binom{n}{e}\binom{n-e}{d}(q-1)^d$ output sequences y with indices e, d. Using this parametric representation, the sets (57), and the definitions of $P_{Y|X}$ in (52) and $Q_Y^\star$ in (53), we obtain

$$\sum_{y \notin \mathcal{S}_{\mathrm{i},x}(c, Q_Y^\star)} P_{Y|X}(y|x) = \sum_{e=0}^{n} \sum_{d=\Psi(e)}^{n-e} \binom{n}{e}\binom{n-e}{d} (q-1)^d \times (1-\delta-\epsilon)^{n-e} \delta^e \varphi^d, \tag{58}$$

and

$$\sum_{y \in \mathcal{S}_{\mathrm{i},x}(c, Q_Y^\star)} Q_Y^\star(y) = \frac{1}{c} \sum_{e=0}^{n} \sum_{d=0}^{\Psi(e)-1} \binom{n}{e}\binom{n-e}{d} (q-1)^d \times (1-\delta-\epsilon)^{n-e} \delta^e \varphi^{\Psi(e)}. \tag{59}$$

Substituting (58) and (59) in (56), reorganizing terms, yields

$$\max_{Q \in \mathcal{Q}} \left\{ \alpha_{\frac{1}{M}} \left(P_X \times P_{Y|X}, P_X \times Q\right)\right\} \geq \sum_{e=0}^{n} \sum_{d=0}^{n-e} \binom{n}{e}\binom{n-e}{d} \times (q-1)^d (1-\delta-\epsilon)^{n-e} \delta^e \varphi^{\max(d,\Psi(e))} - \frac{c}{M}. \tag{60}$$

Finally, noting that

$$\sum_{d=0}^{n-e} \binom{n-e}{d} (q-1)^d = q^{n-e}, \tag{61}$$

we obtain for the normalizing constant in (53),

$$c = \sum_{e=0}^{n} \binom{n}{e} q^{n-e} (1-\delta-\epsilon)^{n-e} \delta^e \varphi^{\Psi(e)}. \tag{62}$$

Substituting (62) in (60), via the meta-converse bound (1), we obtain (55). According to Lemma 3, (55) holds with equality if $\mathcal{C}$ is generalized $Q_Y^\star$-quasi-perfect with parameter $\gamma = c$. ∎

Let $d_{\min}$ denote the minimum Hamming distance between any pair of codewords in $\mathcal{C}$. The Singleton bound [15, Th. 4.5.6] establishes the maximum number of codewords M in a q-ary block code $\mathcal{C}$ of length n and minimum distance $d_{\min}$,

$$\log_q M \leq n - d_{\min} + 1. \tag{63}$$

Those codes achieving the Singleton bound with equality are termed MDS codes. Examples of MDS codes include those that have only two complementary codewords thus having $d_{\min} = n$, non-redundant codes, i.e., $\mathcal{C} = \mathcal{X}$, for which $d_{\min} = 1$, codes with a single parity symbol for which $d_{\min} = 2$, and their corresponding dual codes. These are often called trivial MDS codes. In the case of binary alphabets, only trivial MDS codes exist. For non-binary alphabets, Reed-Solomon codes are the most famous non-trivial MDS codes.

MDS codes are indeed generalized quasi-perfect codes for the q-ary erasure channel ($\epsilon = 0$ in (52)). Then, for any function $\Psi(e) \geq 0$ such that $\Psi(e) = 0$ if, and only if, $e > n - \log_q M$, (53) becomes

$$Q_Y^\star(y) = \begin{cases} 0, & e_y \leq n - \log_q M, \\ \frac{1}{c}\delta^{e_y}(1-\delta)^{n-e_y}, & e_y > n - \log_q M, \end{cases} \tag{64}$$

since (53) abides by the convention $0^0 = 1$.

Consider a generalized $Q_Y^\star$-quasi-perfect code. For the definition of the spheres $\mathcal{S}_x(\cdot)$ we use the convention that, whenever $Q_Y^\star(y) = 0$,

$$\frac{P_{Y|X}(y|x)}{Q_Y^\star(y)} = \begin{cases} 0, & \text{if } P_{Y|X}(y|x) = 0, \\ \infty, & \text{if } P_{Y|X}(y|x) > 0. \end{cases} \tag{65}$$

The spheres induced by this code are such that their interior $\mathcal{S}_{\mathrm{i},x}(c, Q_Y^\star)$ is the set of output sequences y that are compatible with the input x with a number of erasures $e_y \leq n - \log_q M$. Since the codeword-centered interiors do not overlap, the minimum distance of the code is at least $\lfloor n - \log_q M \rfloor + 1$. Since the codeword centered shells $\mathcal{S}_{\mathrm{o},x}(c, Q_Y^\star)$ overlap at some point, then $d_{\min}$ is exactly

$$d_{\min} = \lfloor n - \log_q M \rfloor + 1, \tag{66}$$

which coincides with the Singleton bound (63) when M is a power of q. As a result, we conclude that MDS codes are also quasi-perfect. By letting $\epsilon \to 0$ in Corollary 1 for any $\Psi(e)$ such that $\Psi(e) = 0$ iff $e > n - \log_q M$, we obtain the following result.

Corollary 2: The error probability of any code $\mathcal{C}$ with cardinality M used over a q-ary erasure channel satisfies

$$P_{\mathrm{e}}(\mathcal{C}) \geq \sum_{e=\lfloor n-\log_q M \rfloor+1}^{n} \binom{n}{e} \delta^e (1-\delta)^{n-e} \left(1 - \frac{q^{n-e}}{M}\right), \tag{67}$$

with equality if $\mathcal{C}$ is generalized quasi-perfect with parameters $\gamma = c$ and $Q = Q_Y^\star$, as defined in (64).

The bound in (67) coincides with the converse bound [6, Th. 38]. As observed in [6], this lower bound is tight when $\mathcal{C}$ is an MDS code. Here this result is recovered via the definition of generalized quasi-perfect codes.

We conclude this section with two numerical examples. First, let us consider the transmission of $M = 4$ codewords over a blocklength-n binary input channel (52) for three sets of parameters: BSC with $(\epsilon, \delta) = (0.25, 0)$, a channel with erasures and errors with $(\epsilon, \delta) = (0.05, 0.2)$ and BEC with $(\epsilon, \delta) = (0, 0.25)$. Figure 1 depicts the exact error probability $P_{\mathrm{e}}(\mathcal{C})$ of the best code compared to the lower bound (55) with the choice of $\Psi(e)$ given in (54). The optimal codes for the BSC and BEC are taken from [16] and [17], respectively. For the channel with combined erasures and errors optimal codes are not known for $n \geq 4$ and we use the optimal codes for the BEC from [17], since they offer better performance. Figure 1 shows that the bound (55) for the BSC coincides with the code error probability at the points were quasi-perfect codes exist with respect to the Hamming distance ($n = 2, 3, 4, 5, 6, 8$). For the BEC, the bound (55) (which coincides with (67)) provides the exact error probability at those points where (trivial) MDS codes exist ($n = 2, 3$), as they are generalized quasi-perfect. For the combined errors-erasures channel, to match the lower bound, the codes need to be generalized quasi-perfect for both the BSC and BEC, which only occurs at $n = 2, 3$.

Second, we consider a 32-ary channel (49), and fixed transmission rate $R = \frac{1}{n}\log_{32} M = \frac{1}{2}$. Figure 2 depicts the

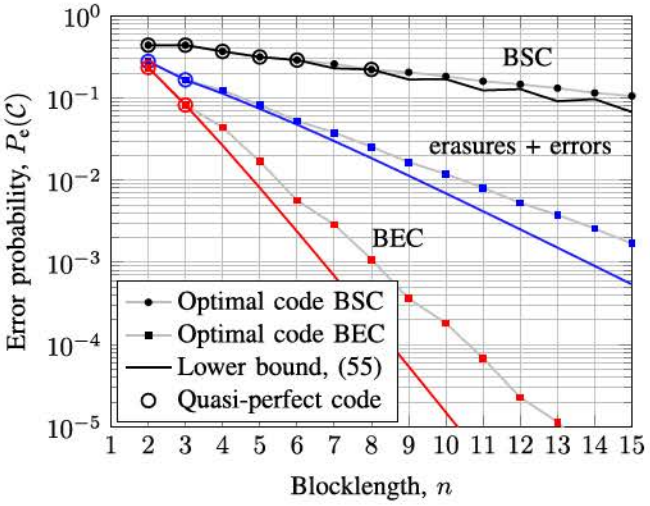

Fig. 1. Error probability for n uses of the channel (52), with $q = 2$, $M = 4$ and BSC: $(\epsilon, \delta) = (0.25, 0)$, erasures and errors: $(\epsilon, \delta) = (0.05, 0.2)$, and BEC: $(\epsilon, \delta) = (0, 0.25)$.

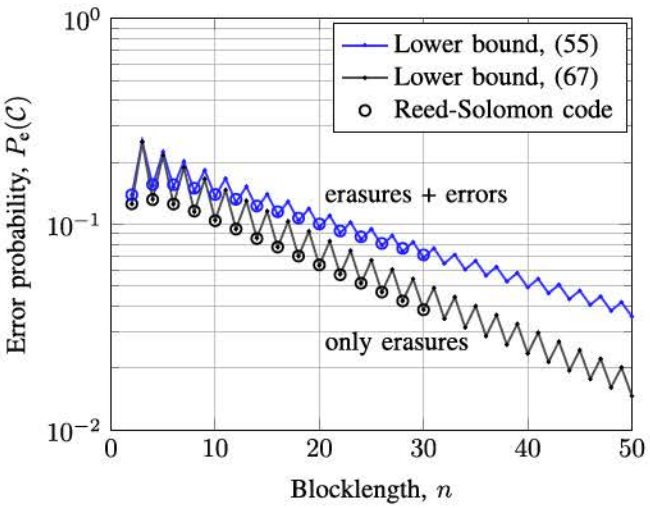

Fig. 2. Error probability for n uses of the channel (52) with $q = 32$, fixed transmission rate $R = \frac{1}{n}\log_{32} M = \frac{1}{2}$, and erasures and errors: $(\epsilon, \delta) = (0.01, 0.36)$, only erasures: $(\epsilon, \delta) = (0, 0.36)$.

lower bound (55) (optimized over a family of functions $\Psi(e)$[5]) for combined erasures and errors with $(\epsilon, \delta) = (0.01, 0.36)$, and the lower bound (67) for erasures only with $(\epsilon, \delta) = (0, 0.36)$. For even blocklengths, we have estimated the performance of the Reed-Solomon code in both scenarios with 10^6 Monte Carlo realizations. Recall that Reed-Solomon codes are defined for blocklengths $n \leq q - 1$ and they are generalized quasi-perfect for the q-ary erasure channel. Therefore, they attain the lower bound (67) with equality in the erasure-only case. While their performance with errors and erasures is not far from the lower bound (55) evaluated with the functions in footnote 5, a gap does exist in this case. Reed-Solomon codes

[5]In particular, the lower bound (55) has been maximized over the functions $\Psi(e)$ of the form $\Psi_0(e) = \max\left(0, \left\lfloor \frac{n-\log_2 M-e+1}{A} \right\rfloor\right)$, $\Psi_1(e) = \max\left(0, \left\lfloor \frac{\lceil n-\log_2 M\rceil-e+1}{A} \right\rfloor\right)$, and $\Psi_2(e) = \max\left(0, \left\lceil \frac{\lfloor n-\log_2 M\rfloor-e+1}{A} \right\rceil\right)$ where $A \in \{1.25, 1.5, 1.75, 2\}$.

can be extended for blocklengths $n = q$ and $n = q + 1$, but there exist no MDS codes for longer blocklengths in general [18].

V. Almost-Lossless Source-Channel Coding

In this section, the notion of quasi-perfect codes is generalized to allow non-equiprobable messages, hence an optimal code needs be matched both to the source and the channel.

We consider the almost-lossless source-channel coding setting. A source generates messages $v \in \mathcal{V}$, where $\mathcal{V}$ is a finite alphabet, according to P_V. The message v is to be transmitted over a channel $P_{Y|X}$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, using a channel encoder that maps each source message v into a codeword $x_v \in \mathcal{X}$. We let $P_{X|V}^{\mathcal{C}}$ denote the conditional distribution $P_{X|V}$ induced by the codebook $\mathcal{C} \triangleq \{x_1, \ldots, x_{|\mathcal{V}|}\}$. The receiver uses maximum-a-posteriori (MAP) decoding to decide on the transmitted message $\hat{v} \in \mathcal{V}$. This decoder minimizes the average error probability, which is given by

$$P_{\mathrm{e}}(\mathcal{C}) = \mathbb{P}\big[\hat{V} \neq V\big] \tag{68}$$

$$= 1 - \sum_{y} \max_{v} P_V(v) P_{Y|X}\big(y|x_v\big). \tag{69}$$

Next, the concept of generalized quasi-perfect codes presented in Section III is further generalized to match both source and channel.

Definition 3: A source-channel code $\mathcal{C}$ is *generalized perfect* with respect to a given source P_V and channel $P_{Y|X}$, if there exists $\tilde{\gamma} \geq 0$ and an auxiliary distribution $Q \in \mathcal{Q}$ such that

$$\bigcup_{v \in \mathcal{V}} \mathcal{S}_{x_v}\left(\frac{\tilde{\gamma}}{P_V(v)}, Q\right) = \mathcal{Y}, \tag{70}$$

where the union is disjoint. More generally, a code is *generalized quasi-perfect* if there exists $\tilde{\gamma} \geq 0$ and $Q \in \mathcal{Q}$ such that (70) is satisfied and the codeword-centered sets $\left\{\mathcal{S}_{\mathrm{i},x_v}\left(\frac{\tilde{\gamma}}{P_V(v)}, Q\right), v \in \mathcal{V}\right\}$ are disjoint.

The definition of a source-channel quasi-perfect code induces a packing of spheres whose radii depend on the probability of the associated source message – more probable source messages are associated to larger spheres. Naturally, if the source messages are equiprobable, then the radii of the spheres become independent of the associated source message and Definition 3 boils down to Definition 2. The generalization of Theorem 1 is as follows.

Theorem 2: Let P_V be the distribution of the source messages, $P_{Y|X}$ be a symmetric channel, and $\mathcal{C}$ be a generalized quasi-perfect source-channel code. Then,

$$P_{\mathrm{e}}(\mathcal{C}) = \min_{P_{X|V}} \max_{Q \in \mathcal{Q}} \left\{ \alpha_{\frac{1}{|\mathcal{V}|}}\left(P_V \times P_{X|V} \times P_{Y|X},\ \bar{P}_V \times P_{X|V} \times Q\right)\right\}, \tag{71}$$

where $\bar{P}_V(v) = \frac{1}{|\mathcal{V}|}$ for all $v \in \mathcal{V}$. Conversely, if (71) holds, then $\mathcal{C}$ is generalized Q-quasi-perfect with respect to the source P_V and channel $P_{Y|X}$.

Proof: See Appendix B. ∎

The right-hand side of (71) is precisely the converse bound [19, Th. 4] particularized to the almost-lossless setting. Therefore, Theorem 2 shows that [19, Th. 4] is tight provided that a generalized quasi-perfect code matched to the source and channel exists.

As a particular case, consider a noiseless channel such that $y = x$ with $\mathcal{X} = \mathcal{Y} = \{1, \ldots, M\}$, and $|\mathcal{V}| > M$. In this case, Definition 3 yields "spheres" of size 1 for the M most probable messages and the $|\mathcal{V}| - M$ least probable messages are assigned to "empty spheres". In practice, the messages associated to these "empty spheres" can be assigned to an arbitrary channel index, as they always yield to a decoding error given their smaller probability. This code corresponds precisely to the well-known optimal almost-lossless block source code. When the M most probable messages have a strictly larger probability than that of the $|\mathcal{V}| - M$ least probable messages, the code is generalized perfect according to Definition 3. When the M-th and $(M+1)$-th most probable messages have the same probability, the code is generalized quasi-perfect.

VI. Lossy Source Coding

In this section, we consider the lossy source coding problem with a maximum distortion constraint. A source generates messages $v \in \mathcal{V}$ with probability distribution P_V. The source encoder maps the message v to a codeword $w \in \mathcal{W}$ belonging to a length-M codebook $\mathcal{C} = \{w_1, w_2, \ldots, w_M\}$. Here $\mathcal{W}$ denotes the reconstruction alphabet. We define a non-negative real-valued distortion measure $d(v, w) : \mathcal{V} \times \mathcal{W} \to \mathbb{R}^+$ and consider a maximum allowed distortion D. The minimum excess-distortion probability of a given code $\mathcal{C}$ is defined as

$$P_{\mathrm{ed}}(\mathcal{C}, D) \triangleq \mathbb{P}\big[d(V, W) > D\big] \tag{72}$$

$$= 1 - \mathbb{P}\Big[\min_{w \in \mathcal{C}} d(V, w) \leq D\Big], \tag{73}$$

where in (73) we used that the minimum excess distortion probability is attained by assigning each source message to the closest (in terms of distortion measure) codeword $w \in \mathcal{C}$.

Quasi-perfect codes have good packing and covering properties simultaneously. Therefore, they are both good channel, as shown in the previous sections, and source codes, as shown next. According to Definition 2 whether a code is generalized quasi-perfect code depends on the channel. In the lossy source-coding setting, this channel turns out to correspond to the test channel induced by the rate-distortion function, although the latter only gives the asymptotic fundamental limit.

Consider a block source encoder that encodes n independent realizations of the source P_V using a codebook of cardinality 2^{nR}. Rate-distortion theory states that, as the blocklength n grows large, the largest rate R of a codebook with maximum distortion D and vanishing excess-distortion probability is given by

$$R(D) \triangleq \min_{P_{W|V} : \mathrm{E}[d(V,W)] \leq D} I(V; W). \tag{74}$$

The optimal $P_{W|V}^{\star}$ in (74) induces a *test channel* $P_{V|W}^{\star}$ that maps the reconstruction points into the source alphabet. More

precisely, let $P^\star_W(w) = \sum_v P_V(v) P^\star_{W|V}(w|v)$, then, Bayes' rule yields $P^\star_{V|W}(v|w) = \frac{P_V(v) P^\star_{W|V}(w|v)}{P^\star_W(w)}$. It is shown in [13, Sec. 10.7] that the optimal test channel has the form

$$P^\star_{V|W}(v|w) = \frac{P_V(v) e^{-\lambda^\star d(v,w)}}{\mu(v)}, \tag{75}$$

for some $\lambda^\star \geq 0$, such that the normalization factor $\mu(v) = \sum_w P^\star_W(w) e^{-\lambda^\star d(v,w)}$ is independent of w.

Let us consider the channel coding problem, as described in Section III, of transmitting M messages over the channel $P^\star_{V|W}$. Good channel codes for $P^\star_{V|W}$ become good source codes for the source P_V and distortion measure $d(v, w)$. In particular, quasi-perfect codes attain the minimum excess-distortion probability, as the next result shows.

Theorem 3: Consider a source P_V with $P_V(v) > 0$, $v \in \mathcal{V}$, distortion measure $d(v, w)$ and maximum distortion D. Let the test channel $P^\star_{V|W}$ in (75) be symmetric and let $\tilde{Q}(v) = \frac{1}{c_\mu} \frac{P_V(v)}{\mu(v)}$ satisfy $\tilde{Q} \in \mathcal{Q}$, where $\mu(v)$ is the normalizing factor in (75) and $c_\mu \triangleq \sum_{v'} \frac{P_V(v')}{\mu(v')}$. Then, the excess-distortion probability of any size-M generalized quasi-perfect code $\mathcal{C}$ with parameters γ and $Q = \tilde{Q}$ is equal to

$$P_{\mathrm{ed}}(\mathcal{C}, D) = \max_{Q_V} \Big\{ \alpha_{M \xi_{\mathcal{C}}(D)}\big(P_V, Q_V\big) \Big\}, \tag{76}$$

where, for any $\mathcal{A} \subseteq \mathcal{W}$,

$$\xi_{\mathcal{A}}(D) \triangleq \sup_{w \in \mathcal{A}} \mathbb{P}\left[d(V, w) \leq D\right], \ V \sim Q_V. \tag{77}$$

Moreover, if $D \geq -\frac{1}{\lambda^\star} \log\big(\gamma / c_\mu\big)$, the excess-distortion probability is $P_{\mathrm{ed}}(\mathcal{C}, D) = 0$.

Proof: See Appendix C. ∎

In [8, Th. 3], the excess-distortion probability of any source code $\mathcal{C}$ (not necessarily quasi-perfect) is expressed as the error probability of an induced binary hypothesis test with certain parameters,

$$P_{\mathrm{ed}}(\mathcal{C}, D) = \max_{Q_V} \Big\{ \alpha_{\rho_{\mathcal{C}}(D)}\big(P_V, Q_V\big) \Big\}, \tag{78}$$

where

$$\rho_{\mathcal{C}}(D) \triangleq \mathbb{P}\Big[\min_{w \in \mathcal{C}} d(V, w) \leq D\Big], \ V \sim Q_V. \tag{79}$$

Invoking

$$\mathbb{P}\Big[\min_{w \in \mathcal{C}} d(V, w) \leq D\Big] \leq M \sup_{w \in \mathcal{C}} \mathbb{P}\left[d(V, w) \leq D\right] \tag{80}$$

$$\leq M \sup_{w \in \mathcal{W}} \mathbb{P}\left[d(V, w) \leq D\right], \tag{81}$$

the identity (78) yields the lower bounds

$$P_{\mathrm{ed}}(\mathcal{C}, D) \geq \max_{Q_V} \Big\{ \alpha_{M \xi_{\mathcal{C}}(D)}\big(P_V, Q_V\big) \Big\} \tag{82}$$

$$\geq \max_{Q_V} \Big\{ \alpha_{M \xi_{\mathcal{W}}(D)}\big(P_V, Q_V\big) \Big\}. \tag{83}$$

Theorem 3 shows that, provided that a quasi-perfect code exists with certain parameters, the lower bound (82) holds with equality. The relaxation from the code to the whole reconstruction alphabet in (83) coincides with [20, Th. 8]. For certain sources, inequality (83) may hold with equality as the next example shows.

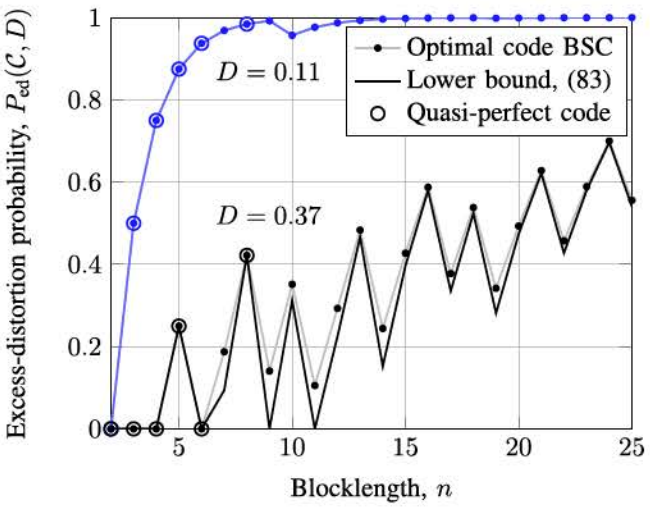

Fig. 3. Minimum excess-distortion probability for n i.i.d. samples of a equiprobable BMS with bit error rate distortion measure and parameters $M = 4$, and $D = 0.11$ (in blue) and $D = 0.37$ (in black).

Let us consider the lossy compression of n i.i.d. samples of an equiprobable binary memoryless source (BMS) with bit error rate distortion measure, i.e., $P_V(v) = 2^{-n}$ and $d(v, w) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}[v_i \neq w_i]$, with $v, w \in \{0, 1\}^n$. The test channel for this rate-distortion problem corresponds to a BSC with a crossover probability depending on D. As in the channel coding example from Fig. 1, we consider a codebook with $M = 4$ codewords. Figure 3 depicts the minimum excess-distortion probability $P_{\mathrm{ed}}(\mathcal{C}, D)$ as a function of n for a maximum distortion $D = 0.11$ and $D = 0.37$. Since we are "quantizing" a space of increasing dimension n with only $M = 4$ codewords, the excess-distortion probability tends to 1 as $n \to \infty$ for any $D < \frac{1}{2}$. In Fig. 3, we plot the lower bound (83) evaluated for Q_V uniform [20, Th. 15], compared to the exact excess-distortion probability evaluated for the best code in a BSC and $M = 4$ codewords [16]. We also highlight with markers the points where quasi-perfect codes exist for the BSC, corresponding to $n = 2, 3, 4, 5, 6, 8$ (see Fig. 1).

In Fig. 3, we observe that the exact excess-distortion probability coincides with the lower bound (83) at the points where quasi-perfect codes exist both for $D = 0.11$ and $D = 0.37$. Nevertheless, in the lossy compression setting, the reverse implication is not always true. Depending on the system parameters, the exact excess-distortion probability and the lower bound can also coincide even when no quasi-perfect code exists for the corresponding test channel. Indeed, for $D = 0.37$ the only points where the exact excess-distortion probability and the lower bound coincide are those for which quasi-perfect codes exist, while for $D = 0.11$ both expressions coincide for all values of n, regardless of whether the code is quasi-perfect. This occurs when the sets $\{v \in \mathcal{V} \,|\, d(V, w) \leq D\}$, $w \in \mathcal{C}$, are non-overlapping (this occurs in our example for D sufficiently small). Then, the encoding regions which satisfy the maximum distortion cap are "spheres" regardless of the specific structure of the codebook $\mathcal{C}$ and the lower bound (83) yields the exact excess-distortion probability.

VII. Discussion

We have proposed a generalization of perfect and quasi-perfect codes beyond the Hamming distance and their conventional application to binary symmetric channels. The definition of these codes follows from the packing and covering properties of a set of generalized spheres whose shape is tilted using an auxiliary probability measure on the output alphabet. Since the shape of these spheres depends on the channel considered, quasi-perfect codes can only be defined with respect to a specific channel. For the BSC, quasi-perfect codes are defined with respect to the Hamming distance and our definition recovers the classical definition of quasi-perfect (or sphere-packed) codes in the literature. The tilting of these spheres with a cleverly chosen auxiliary measure shows that MDS codes are quasi-perfect for erasure channels. The key property satisfied by the generalized quasi-perfect codes is that they achieve the minimum error probability for a given blocklength and rate.

While the proofs of the results in this paper are presented for discrete channels, they can be extended for channels with continuous outputs with some care. In fact, Lemma 1, Definitions 1 and 2, and Theorem 1 apply without change for both discrete and continuous channels, provided that they are absolutely continuous with respect to Lebesgue measure. Nevertheless, the spheres induced by typical continuous channels seldom allow a perfect (or quasi-perfect) packing of the output space. Some atypical examples of continuous channels in which the induced spheres pack the output space are the additive white Gaussian noise (AWGN) channel with $M = 2$ codewords with equal power, the binary-input AWGN when all the input sequences are used, i.e., for $M = 2^n$, or the additive white Laplace noise channel under certain input constraints (as the induced spheres are norm-1 balls, and thus they can pack the space for specific lattice codes).

The framework presented in this work has been built upon the key assumption that a certain natural channel symmetry holds. Nevertheless, the underlying idea can be applied to general channels $P_{Y|X}$ and arbitrary auxiliary distributions Q. In this case, quasi-perfect codes are defined as those "codes attaining the meta-converse bound with equality." This definition is reminiscent to that of the MDS codes, which are defined as "codes attaining the Singleton bound with equality." Admittedly, while this alternative more general definition of quasi-perfect codes is mathematically precise, it does not shed much light into the structure of these codes.

Appendix A
Proof of Lemma 1

For any function $A(z)$ and any test $0 \leq T(z) \leq 1$, the following simple inequality holds

$$\sum_{z\in\mathcal{Z}} A(z)\mathbb{1}\big[A(z) > 0\big] \geq \sum_{z\in\mathcal{Z}} A(z)T(z). \tag{84}$$

Particularized with $A(z) = P_0(z) - \gamma\, P_1(z)$, $\gamma \geq 0$, and $T = T_{\text{NP}}$ in (5), the inequality (84) yields

$$\sum\nolimits_z \big(P_0(z) - \gamma\, P_1(z)\big)\mathbb{1}\big[P_0(z) - \gamma\, P_1(z) > 0\big] \geq \sum\nolimits_z \big(P_0(z) - \gamma\, P_1(z)\big)T_{\text{NP}}(z). \tag{85}$$

Rearranging terms in (85), we obtain

$$-\sum\nolimits_z P_0(z)T_{\text{NP}}(z) \geq -\sum\nolimits_z P_0(z)\mathbb{1}\left[\frac{P_0(z)}{P_1(z)} > \gamma\right] + \gamma \sum\nolimits_z P_1(z)\left(\mathbb{1}\left[\frac{P_0(z)}{P_1(z)} > \gamma\right] - T_{\text{NP}}(z)\right). \tag{86}$$

Adding 1 to both sides of (86) and noting that $\alpha_\beta\big(P_0, P_1\big) = 1 - \sum_z P_0(z)T_{\text{NP}}(z)$ for $\beta = \sum_z P_1(z)T_{\text{NP}}(z)$, yields

$$\alpha_\beta\big(P_0, P_1\big) \geq \sum\nolimits_z P_0(z)\mathbb{1}\left[\frac{P_0(z)}{P_1(z)} \leq \gamma\right] + \gamma \sum\nolimits_z P_1(z)\mathbb{1}\left[\frac{P_0(z)}{P_1(z)} > \gamma\right] - \gamma\beta, \tag{87}$$

which coincides with the right-hand side of (6) for fixed γ.

We now show that (87) holds with equality when γ coincides with the threshold appearing in the Neyman-Pearson test (5). To see this, note that second indicator function in (5) is active only when $P_0(z) - \gamma\, P_1(z) = 0$, and equal to 0 otherwise. Then, multiplying both sides of (5) by $P_0(z) - \gamma\, P_1(z)$, summing over z, yields (85) with equality. Since (85) holds with equality for γ equal to the threshold appearing in the Neyman-Pearson test, so it does (87). Then, by optimizing (87) over thresholds $\gamma \geq 0$, we obtain the equality in (6) and the result follows.

Appendix B
Proof of Theorem 2

We apply Lemma 1 to the hypothesis test in (71) to obtain an alternative expression for the Neyman-Paerson performance of the test. This expression is then shown to coincide with the following characterization of the joint source-channel error probability of a quasi-perfect code.

For a source-channel code, we define the following countable set, which is analogous to $\mathcal{L}_Q$ in (15),

$$\mathcal{L}_Q^{(v)} \triangleq \left\{\tilde{\nu} \in \mathbb{R} \,\middle|\, \exists x \in \mathcal{X}, \exists y \in \mathcal{Y}, \tfrac{P_{Y|X}(y|x)}{Q(y)} = \tfrac{\tilde{\nu}}{P_V(v)}\right\}. \tag{88}$$

Lemma 4: For a source P_V and a symmetric channel $P_{Y|X}$, the error probability of any source-channel code $\mathcal{C}$ satisfies, for any $\tilde{\gamma} \geq 0$ and any $Q \in \mathcal{Q}$,

$$P_{\text{e}}(\mathcal{C}) \geq \tilde{\gamma}\left(\sum_v \mathsf{Q}_{\text{i}}\left(\tfrac{\tilde{\gamma}}{P_V(v)}\right) - 1\right) + \sum_v \sum_{\tilde{\tau}\in\mathcal{L}_Q^{(v)}\cap[0,\tilde{\gamma}]} \tilde{\tau}\mathsf{Q}_{\text{o}}\left(\tfrac{\tilde{\tau}}{P_V(v)}\right). \tag{89}$$

Furthermore, the lower bound (89) holds with equality if and only if the source-channel code $\mathcal{C}$ is generalized quasi-perfect with (not necessarily unique) parameters $\tilde{\gamma}$ and Q satisfying the conditions in Definition 3.

Proof: The proof follows analogous steps to that of Lemma 3, and it is omitted here. Indeed, for $|\mathcal{V}| = M$ and $P_V(v) = \frac{1}{M}$, letting $\tilde{\gamma} = \frac{\gamma}{M}$, $\tilde{\tau} = \frac{\tau}{M}$, then (89) recovers the right-hand side of (22), which is tight for quasi-perfect codes satisfying Definition 2 with parameters γ and Q. ∎

Applying Lemma 1 with $P_0 \leftarrow P_V P_{X|V} P_{Y|X}$ and $P_1 \leftarrow \bar{P}_V P_{X|V} Q$, via the change of variable $\gamma \leftrightarrow \tilde{\gamma} = \frac{\gamma}{|\mathcal{V}|}$, yields

$$\begin{aligned}
&\alpha_{\frac{1}{|\mathcal{V}|}}\Big(P_V \times P_{X|V} \times P_{Y|X},\ \bar{P}_V \times P_{X|V} \times Q\Big) \\
&= \max_{\tilde{\gamma}\geq 0}\Bigg\{\sum_{v,x} P_V(v) P_{X|V}(x|v) \sum_{y\notin \mathcal{S}_{\mathrm{i},x}\left(\frac{\tilde{\gamma}}{P_V(v)}, Q\right)} P_{Y|X}(y|x) \\
&\qquad + \tilde{\gamma}\sum_{v,x} P_{X|V}(x|v) \sum_{y\in \mathcal{S}_{\mathrm{i},x}\left(\frac{\tilde{\gamma}}{P_V(v)}, Q\right)} Q(y) - \tilde{\gamma}\Bigg\} \qquad (90) \\
&= \max_{\tilde{\gamma}\geq 0}\Bigg\{\sum_{v,x} P_V(v) P_{X|V}(x|v) \sum_{\tilde{\tau}\in \mathcal{L}_Q^{(v)}\cap[0,\tilde{\gamma}]} \frac{\tilde{\tau}}{P_V(v)} \mathsf{Q}_{\mathrm{o}}\left(\frac{\tilde{\tau}}{P_V(v)}\right) \\
&\qquad + \tilde{\gamma}\sum_{v,x} P_{X|V}(x|v)\, \mathsf{Q}_{\mathrm{i}}\left(\frac{\tilde{\gamma}}{P_V(v)}\right) - \tilde{\gamma}\Bigg\}, \qquad (91)
\end{aligned}$$

where in the last step we used that the complementary set of $\mathcal{S}_{\mathrm{i},x}\left(\frac{\tilde{\gamma}}{P_V(v)}, Q\right)$ corresponds to $\bigcup_{\tilde{\tau}\in\mathcal{L}_Q^{(v)}\cap[0,\tilde{\gamma}]} \mathcal{S}_{\mathrm{o},x}\left(\frac{\tilde{\tau}}{P_V(v)}, Q\right)$, that $\frac{P_{Y|X}(y|x)}{Q(y)} = \frac{\tilde{\tau}}{P_V(v)}$ for all $y \in \mathcal{S}_{\mathrm{o},x}\left(\frac{\tilde{\tau}}{P_V(v)}, Q\right)$. Finally, using that $\sum_x P_{X|V}(x|v) = 1$, we obtain

$$\begin{aligned}
&\alpha_{\frac{1}{|\mathcal{V}|}}\Big(P_V \times P_{X|V} \times P_{Y|X},\ \bar{P}_V \times P_{X|V} \times Q\Big) \\
&\qquad = \max_{\tilde{\gamma}\geq 0}\Bigg\{\sum_{v} \sum_{\tilde{\tau}\in \mathcal{L}_Q^{(v)}\cap[0,\tilde{\gamma}]} \tilde{\tau}\mathsf{Q}_{\mathrm{o}}\left(\frac{\tilde{\tau}}{P_V(v)}\right) \\
&\qquad\qquad + \tilde{\gamma}\sum_{v} \mathsf{Q}_{\mathrm{i}}\left(\frac{\tilde{\gamma}}{P_V(v)}\right) - \tilde{\gamma}\Bigg\}, \qquad (92)
\end{aligned}$$

which coincides with (89) when $\tilde{\gamma}$ coincides with its optimizing value in (92). Since (92) is a lower bound to $P_{\mathrm{e}}(\mathcal{C})$, the theorem thus follows by optimizing (92) over auxiliary distributions $Q \in \mathcal{Q}$.

Appendix C
Proof of Theorem 3

Let $\mathcal{C}$ be generalized quasi-perfect with respect to the test channel $P^{\star}_{V|W}$ defined in (75), with parameters γ and $\tilde{Q}(v) = \frac{1}{c_\mu}\frac{P_V(v)}{\mu(v)}$. The set $\mathcal{S}_w(\tau, \tilde{Q})$ associated to the test channel $P^{\star}_{W|V}$ is given by

$$\mathcal{S}_w(\tau, \tilde{Q}) = \left\{v \in \mathcal{V} \,\middle|\, d(v,w) \leq -\frac{1}{\lambda^\star}\log\left(\tau\frac{\mu(v)\tilde{Q}(v)}{P_V(v)}\right)\right\}, \qquad (93)$$

which upon particularization to $\tilde{Q}(v) = \frac{1}{c_\mu}\frac{P_V(v)}{\mu(v)}$ yields

$$\mathcal{S}_w(\tau, \tilde{Q}) = \left\{v \in \mathcal{V} \,\middle|\, d(v,w) \leq -\frac{1}{\lambda^\star}\log\frac{\tau}{c_\mu}\right\}. \qquad (94)$$

We divide the proof in two different cases depending on the value of the maximum distortion D.

a) $D \geq -\frac{1}{\lambda^\star}\log(\gamma/c_\mu)$: In this case $\gamma \geq c_\mu e^{-\lambda^\star D}$, and

$$\mathcal{S}_w(\gamma, \tilde{Q}) \subseteq \mathcal{S}_w\big(c_\mu e^{-\lambda^\star D}, \tilde{Q}\big) = \Big\{v \in \mathcal{V} \,\Big|\, d(v,w) \leq D\Big\}. \qquad (95)$$

According to Definition 2, the codeword-centered sets $\big\{\mathcal{S}_w(\gamma, \tilde{Q})\big\}_{w\in\mathcal{C}}$ cover the space. Then, using (95) it follows that every element of $\mathcal{V}$ has a codeword no farther than D, i.e.,

$$\bigcup_{w\in\mathcal{C}}\Big\{v \in \mathcal{V} \,\Big|\, d(v,w) \leq D\Big\} = \mathcal{V}. \qquad (96)$$

According to (96), we have that $\mathbb{P}\big[\min_{w\in\mathcal{C}} d(V,w) \leq D\big] = 1$ regardless of the distribution of V. As a result, the excess-distortion probability is

$$P_{\mathrm{ed}}(\mathcal{C}, D) = 1 - \mathbb{P}\Big[\min_{w\in\mathcal{C}} d(v,w) \leq D\Big] = 0. \qquad (97)$$

Similarly, for $\xi_{\mathcal{C}}(D)$ and $\rho_{\mathcal{C}}(D)$ defined in (77) and (79), using (81) it follows that $M\xi_{\mathcal{C}}(D) \geq \rho_{\mathcal{C}}(D) = 1$. Since $\alpha_1\big(P_V, Q_V\big) = 0$, using (97), we conclude that (76) holds with equality.

b) $D < -\frac{1}{\lambda^\star}\log(\gamma/c_\mu)$: In this region, $\gamma < c_\mu e^{-\lambda^\star D}$, and it thus follows that

$$\mathcal{S}_{\mathrm{i},w}(\gamma, \tilde{Q}) \supseteq \mathcal{S}_w\big(c_\mu e^{-\lambda^\star D}, \tilde{Q}\big) = \Big\{v \in \mathcal{V} \,\Big|\, d(v,w) \leq D\Big\}. \qquad (98)$$

In this case, $\bigcup_{w\in\mathcal{C}}\big\{v \in \mathcal{V} \,\big|\, d(v,w) \leq D\big\}$ does not cover the space completely. Nevertheless, since the code $\mathcal{C}$ is quasi-perfect with radius γ, the spheres $\mathcal{S}_{\mathrm{i},w}(\gamma, Q)$, $w \in \mathcal{C}$, are disjoint. Using (98) we conclude that the sets $\big\{v \in \mathcal{V} \,\big|\, d(v,w) \leq D\big\}$, $w \in \mathcal{C}$, do not overlap. Therefore,

$$\begin{aligned}
P_{\mathrm{ed}}(\mathcal{C}, D) &= 1 - \mathbb{P}\Big[\min_{w\in\mathcal{C}} d(V,w) \leq D\Big] \qquad (99) \\
&= 1 - \mathbb{P}\Big[v \in \bigcup_{w\in\mathcal{C}}\Big\{v \in \mathcal{V} \,\Big|\, d(v,w) \leq D\Big\}\Big] \qquad (100) \\
&= 1 - \sum_{w\in\mathcal{C}} \mathbb{P}\big[d(V,w) \leq D\big], \qquad (101)
\end{aligned}$$

where $V \sim P_V$.

We now show that the right-hand sides of (76) and (101) coincide. Applying Lemma 1 to the hypothesis test in (76), yields

$$\begin{aligned}
\alpha_\beta\big(P_V, Q_V\big) = \max_{\gamma'\geq 0}\Bigg\{&\mathbb{P}\left[\frac{P_V(V)}{Q_V(V)} \leq \gamma'\right] \\
&+ \gamma'\mathbb{P}\left[\frac{P_V(\bar{V})}{Q_V(\bar{V})} > \gamma'\right] - \gamma'\beta\Bigg\}, \qquad (102)
\end{aligned}$$

where $V \sim P_V$ and $\bar{V} \sim Q_V$. Let

$$Q_V^{\mathcal{C}}(v) \triangleq \frac{1}{g}P_V(v)\left(\frac{1}{M}\sum_{w\in\mathcal{C}} e^{-\lambda d(v,w)}\right)^{-1} \qquad (103)$$

where g is a normalizing factor and $\lambda \geq 0$ is to be defined later. Using $Q_V = Q_V^{\mathcal{C}}$ and choosing $\gamma' = \frac{g}{M}e^{-\lambda D}$, we obtain the following lower bound to (102),

$$\alpha_\beta\big(P_V, Q_V^{\mathcal{C}}\big) \geq \mathbb{P}\left[\sum_{w\in\mathcal{C}} e^{-\lambda d(V,w)} \leq e^{-\lambda D}\right]$$

$$+\frac{g}{M}e^{-\lambda D}\left(\mathbb{P}\Big[\sum_{w\in\mathcal{C}} e^{-\lambda d(\bar{V},w)} > e^{-\lambda D}\Big] - \beta\right), \quad (104)$$

where $V \sim P_V$ and $\bar{V} \sim Q_V^{\mathcal{C}}$.

For $\lambda \geq 0$ sufficiently large,

$$\sum_{w\in\mathcal{C}} e^{-\lambda d(v,w)} > e^{-\lambda D} \Leftrightarrow \min_{w\in\mathcal{C}} d(v,w) \leq D. \quad (105)$$

Therefore, for such λ, (104) becomes

$$\alpha_\beta\big(P_V, Q_V^{\mathcal{C}}\big) \geq \mathbb{P}\Big[\min_{w\in\mathcal{C}} d(V,w) > D\Big] + \frac{g}{M}e^{-\lambda D}\left(\mathbb{P}\Big[\min_{w\in\mathcal{C}} d(\bar{V},w) \leq D\Big] - \beta\right). \quad (106)$$

The symmetry conditions required by the theorem imply that the measure of the set $\{v \in \mathcal{V} \,|\, d(v,w) \leq \delta\}$ does not depend on $w \in \mathcal{W}$ for any $\delta \geq 0$. Then, since the sets $\{v \in \mathcal{V} \,|\, d(v,w) \leq \delta\}$ are non-overlapping, for sufficiently large λ, we obtain

$$\mathbb{P}\Big[\min_{w\in\mathcal{C}} d(\bar{V},w) \leq D\Big] = \sum_{w\in\mathcal{C}} \mathbb{P}\big[d(\bar{V},w) \leq D\big] \quad (107)$$

$$= M \sup_{w\in\mathcal{C}} \mathbb{P}\left[d(\bar{V},w) \leq D\right], \quad (108)$$

where in (108) we used that, for sufficiently large λ, $Q_V^{\mathcal{C}}(v)$ only depends on the distance to the closest $w \in \mathcal{C}$. Then, since the measure of the set $\{v \in \mathcal{V} \,|\, d(v,w) = \delta\}$ does not depend on $w \in \mathcal{W}$ for any $\delta \geq 0$, neither does $\mathbb{P}\left[d(\bar{V},w) = \delta\right]$ nor $\mathbb{P}\left[d(\bar{V},w) \leq D\right]$ depend on $w \in \mathcal{C}$.

Therefore, for $\beta = \xi_{\mathcal{C}}(D)$, (106) becomes

$$\alpha_{M\xi_{\mathcal{C}}(D)}\big(P_V, Q_V\big) \geq 1 - \sum_{w\in\mathcal{C}} \mathbb{P}\big[d(V,w) \leq D\big]. \quad (109)$$

Since the left-hand side of (109) is a lower bound to $P_{\text{ed}}(\mathcal{C}, D)$, and since the right-hand side of (109) coincides with (101), we conclude that (76) holds with equality.

Remark: Note that $\mathbb{P}\big[d(\bar{V},w) \leq D\big]$, $\bar{V} \sim Q_V^{\mathcal{C}}$, becomes independent of $w \in \mathcal{C}$ as $\lambda \to \infty$. However, for this choice of $\bar{V}$, the measure $\mathbb{P}\big[d(\bar{V},w) \leq D\big]$ still depends on $w \notin \mathcal{C}$. Therefore, the proof technique presented here cannot be directly applied when the β parameter in (76) is relaxed from $M\xi_{\mathcal{C}}(D)$ to $M\xi_{\mathcal{W}}(D)$, as discussed in (82)-(83).

References

[1] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Inf. Control*, vol. 10, no. 1, pp. 65–103, Jan. 1967.

[2] E. A. Haroutuninan, "Estimates of the error exponents for the semi-continuous memoryless channel," (in Russian) *Problemy Peredachi Inf.*, vol. 4, no. 4, pp. 37–48, 1968.

[3] R. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 405–417, Jul. 1974.

[4] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1962–1990, May 2008.

[5] Y. Altuğ and A. B. Wagner, "Refinement of the sphere-packing bound: Asymmetric channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1592–1614, Mar. 2014.

[6] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[7] R. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[8] G. Vazquez-Vilar, A. Tauste Campo A. Guillén i Fàbregas, and A. Martinez, "Bayesian M-Ary hypothesis testing: The meta-converse and Verdú-Han bounds are tight," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2324–2333, May 2016.

[9] M. Hamada, "A sufficient condition for a code to achieve the minimum decoding error probability–generalization of perfect and quasi-perfect codes," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E83-A, no. 10, pp. 1870–1877, Oct. 2000.

[10] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication*. Cambridge, MA, USA: Massachusetts Institute of Technology, 1961.

[11] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London A, Math. Phys. Sci.*, vol. 231, pp. 289–337, Jan. 1933.

[12] J. L. Massey. (1998). *Applied Digital Information Theory I, Lecture Notes*. [Online]. Available: http://www.isi.ee.ethz.ch/research.html

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[14] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.

[15] S. Roman, *Coding and Information Theory*. New York, NY, USA: Springer, 1992.

[16] P.-N. Chen, H.-Y. Lin, and S. M. Moser, "Optimal ultrasmall block-codes for binary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7346–7378, Nov. 2013.

[17] H.-Y. Lin, S. M. Moser, and P.-N. Chen, "Weak flip codes and their optimality on the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5191–5218, Jul. 2018.

[18] G. Seroussi and R. M. Roth, "On MDS extensions of generalized Reed-Solomon codes," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 3, pp. 349–354, May 1986.

[19] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[20] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.

Gonzalo Vazquez-Vilar (S'08–M'12) received the Telecommunication Engineering degree from the University of Vigo, Spain, in 2004, the Master of Science degree from Stanford University, U.S., in 2008 and the Ph.D. in Communication Systems from the University of Vigo, Spain, in 2011.

In 2011–2014 he was a post-doctoral fellow in the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Spain and since 2014 he has been a Visiting Professor in the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Spain. He has held appointments as visiting researcher at Stanford University, U.S., University of Cambridge, U.K., and Princeton University, U.S. His research interests lie in the field of Shannon theory, with emphasis on finite-length information theory and communications.

Albert Guillén i Fàbregas (S'01–M'05–SM'09) received the Telecommunication Engineering degree and the Electronics Engineering degree from the Universitat Politècnica de Catalunya, and the Politecnico di Torino, Torino, Italy, respectively, in 1999, and the Ph.D. degree in Communication Systems from Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2004.

Since 2011 he has been an ICREA Research Professor at Universitat Pompeu Fabra. He is also an Adjunct Researcher at the University of Cambridge. He has held appointments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, University of Cambridge, as well as visiting appointments at EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra, University of South Australia, Centrum Wiskunde & Informatica and Texas A&M University in Qatar. His research interests are in the areas of information theory, coding theory and communication theory.

Dr. Guillén i Fàbregas is a member of the Young Academy of Europe, received both Starting and Consolidator Grants from the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference (EUSIPCO), the 2004 Nokia Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a pre-doctoral Research Fellowship of the Spanish Ministry of Education to join ESA. He was a general co-chair of the 2016 IEEE International Symposium on Information Theory held in Barcelona and a technical-program committee co-chair of the 2013 IEEE Information Theory Workshop held in Sevilla. He is an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY, an Editor of the *Foundations and Trends in Communications and Information Theory* and was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Sergio Verdú (S'80–M'84–SM'88–F'93) received the Telecommunications Engineering degree from the Universitat Politècnica de Barcelona in 1980, and the Ph.D. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign in 1984. He was on the faculty of Princeton University from 1984 to 2018.

Sergio Verdú is the recipient of the 2007 Claude E. Shannon Award, and the 2008 IEEE Richard W. Hamming Medal. He is a member of both the National Academy of Engineering and the National Academy of Sciences. In 2016, Verdú received the National Academy of Sciences Award for Scientific Reviewing.

Verdú is a recipient of several paper awards from the IEEE: the 1992 Donald Fink Paper Award, the 1998 and 2012 Information Theory Paper Awards, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Prize Award, the 2006 Joint Communications/Information Theory Paper Award, and the 2009 Stephen O. Rice Prize from the IEEE Communications Society. In 1998, Cambridge University Press published his book *Multiuser Detection,* for which he received the 2000 Frederick E. Terman Award from the American Society for Engineering Education. He was awarded a Doctorate Honoris Causa from the Universitat Politècnica de Catalunya in 2005, and was elected corresponding member of the Real Academia de Ingeniería of Spain in 2013.

Sergio Verdú served as President of the IEEE Information Theory Society in 1997, and on its Board of Governors (1988-1999, 2009-2014). He has also served in various editorial capacities for the IEEE TRANSACTIONS ON INFORMATION THEORY: Associate Editor (Shannon Theory, 1990-1993; Book Reviews, 2002-2006), Guest Editor of the Special *50th Anniversary Commemorative Issue* (published by IEEE Press as Information Theory: Fifty years of discovery), and member of the Executive Editorial Board (2010-2013). He served as Creative Producer of the film The Bit Player, a documentary on Claude Shannon's life and legacy. He co-chaired the Europe-United States *Frontiers of Engineering* program, of the National Academy of Engineering during 2009-2013. He served as the founding Editor-in-Chief of *Foundations and Trends in Communications and Information Theory*. Verdú served as co-chair of the 2000 and 2016 *IEEE International Symposia on Information Theory*.

Sergio Verdú has held visiting appointments at the Australian National University, the Technion-Israel Institute of Technology, the University of Tokyo, the University of California, Berkeley, the Mathematical Sciences Research Institute, Berkeley, Stanford University, and the Massachusetts Institute of Technology.